Whitepaper

Disaggregated Spine for Modern AI Networks

Introduction

Data center (DC) spine layers have historically been implemented using large, multi-terabit chassis-based systems. These systems consist of multiple line cards interconnected through a proprietary, centralized switching fabric, forming a vertically scaled, monolithic platform. While chassis-based systems provide the necessary aggregate spine radix to support both intra–DC traffic and Data Center Interconnect (DCI) requirements, the unified architecture does not permit independent optimization for these distinct traffic domains.

The rapid growth of accelerator (xPU) deployments further highlights this limitation. Modern AI workloads increasingly rely on scale-across traffic models in which training and inference span multiple geographically distributed data centers. As a result, inter–data center traffic volumes are growing at rates that approach, and in some cases rival, intra–data center traffic.

This shift increases divergence between inter–data center and intra–data center requirements, reinforcing the need for architectural separation within the spine layer.

Disaggregated Spine Architecture

A disaggregated spine architecture addresses this limitation by decomposing the monolithic spine layer into independent functional tiers. One tier aligns with Data Center Interconnect (DCI) requirements, including encrypted egress traffic, long-haul optical transport, expanded routing scale, and deeper buffering to accommodate wide-area traffic. The other tier aligns with intra–data center traffic, emphasizing high port density, low latency, and efficient aggregation within the data center fabric.

The DCI-facing tier is defined as the **Scale-across Spine,** supporting traffic that spans geographically distributed data centers. The fabric-facing tier is defined as the **Scale-across Leaf**, supporting large-scale intra–data center traffic aggregation.

Mar 3, 2026

Separating Scale-across Spine and Scale-across Leaf functions into distinct architectural layers enables independent optimization according to the operational and performance requirements of each traffic domain, rather than satisfying both within a single integrated chassis (see Figure 1).

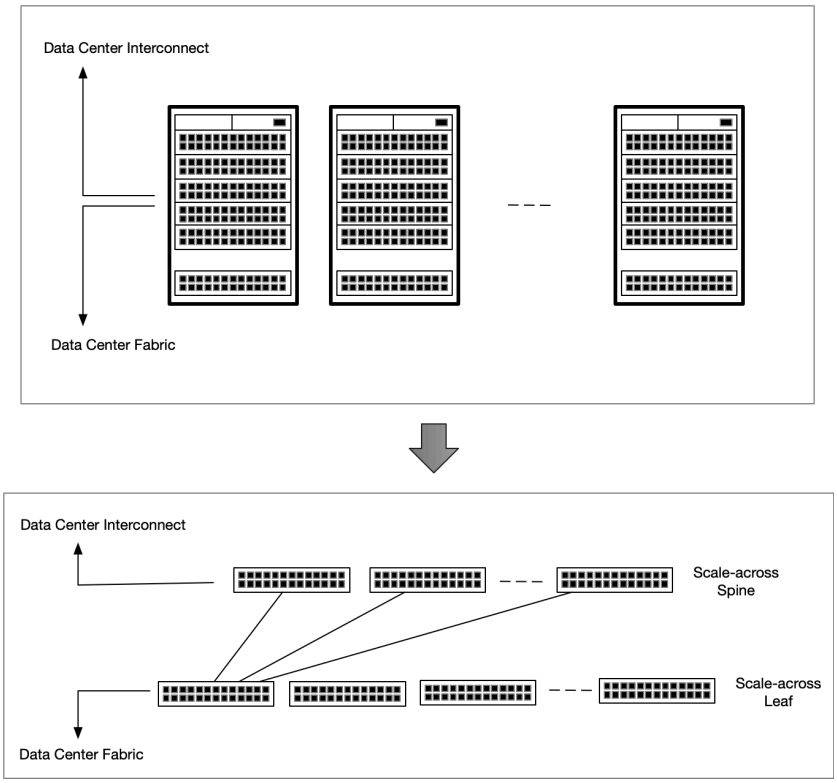

Figure 1: From Chassis-based Spine to Scale-across Spine and Scale-across Leaf

Architectural Benefits of Disaggregation

The architectural separation of Scale-across Spine and Scale-across Leaf functions provides three primary advantages: design flexibility, operational efficiency, and cost and power efficiency.

1. Design Flexibility

By decomposing the spine layer into independent tiers, each can be optimized according to its specific traffic characteristics rather than satisfying the combined requirements of DCI and intra–data center workloads within a single chassis. Silicon selection, buffering

strategy, optical interface and transceiver technology, as well as feature depth, can evolve independently across tiers. This separation establishes distinct upgrade domains within the spine layer, reducing the scope of change required to introduce new silicon generations or within-generation optimizations. The current scale-across spine supports either 400 GbE or 800 GbE deployments and provides a migration path to 1.6 Tb/s-class silicon in subsequent generations without requiring wholesale spine replacement. Furthermore, this scale out architecture allows for an incremental capacity planning for the scale-across networks that need to handle bandwidth and footprint growth in short order.

2. Operational Efficiency

Functional decoupling reduces system-wide coordination during upgrades and scaling events. Independent spine tiers allow targeted expansion, maintenance, and troubleshooting without affecting unrelated traffic domains. The use of standards-based Ethernet interconnects further improves visibility and operational consistency compared to proprietary fabrics.

3. Cost and Power Efficiency

Architectural separation also has direct economic implications. In a disaggregated spine, each tier can be implemented using switch silicon aligned with its specific functional requirements rather than selecting a single silicon class capable of satisfying both traffic domains.

The interconnect-facing tier requires encryption support, expanded routing scale, long-haul optical interfaces, and deeper buffering. These capabilities necessitate feature-rich silicon—including encryption engines, larger routing tables, and deeper packet buffers—with corresponding cost and power characteristics.

In contrast, the fabric-facing tier does not require encryption, large route tables, or deep buffering. It is therefore well suited to high-density switching silicon optimized for port scale and latency performance. Such silicon is typically offered at approximately twice the port density and at roughly half the cost per device relative to silicon designed for interconnect-facing roles.

By selecting silicon independently for each tier, the disaggregated spine avoids over-provisioning feature depth where it is unnecessary and improves capital efficiency and power-per-port characteristics.

In configurations delivering an equivalent number of service ports and identical oversubscription ratios, empirical evaluation demonstrates approximately 30% reductions in both system cost and total power consumption relative to chassis-based designs. These savings provide additional design flexibility: operators may increase aggregate capacity or reduce oversubscription ratios while maintaining comparable power and cost envelopes.

Interconnect technology selection further influences efficiency. Comparative evaluation of fully retimed optical transceivers and linear pluggable optics (LPO) demonstrates additional reductions in power consumption when lower-power interconnect technologies are deployed.

The approximately 30% savings are observed when 51.2 Tb/s-class silicon is deployed in the fabric-facing tier. With 102.4 Tb/s-class silicon at this tier, total power reductions approach 40%, as shown in Table 1.

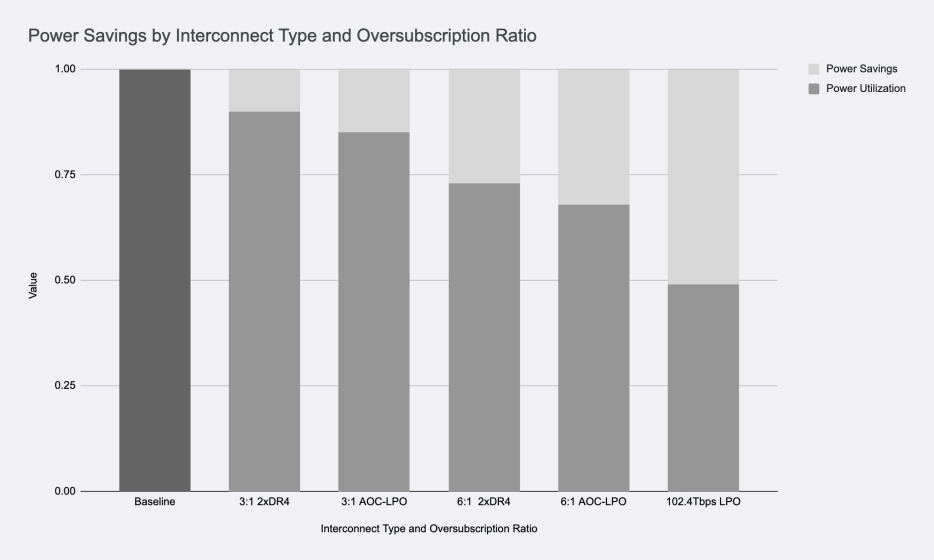

Table 1: Power savings achieved by the Disaggregated Spine across interconnect types and oversubscription ratios.

Open Software Architecture and Standards

The architectural decomposition of the spine layer not only enables hardware-level separation of functional domains, but also facilitates adoption of open network operating systems such as SONiC. In contrast to chassis-based systems, where the operating system is tightly integrated with proprietary hardware subsystems and centralized fabric control,

each spine node in a disaggregated architecture operates as an independent switching element.

Additionally, the interconnect between fabric-facing and interconnect-facing spine nodes is implemented using standards-based Ethernet rather than a proprietary internal switching fabric. This design choice simplifies interoperability, enhances operational visibility, and improves troubleshooting by leveraging widely understood Ethernet tools and telemetry mechanisms. By eliminating proprietary fabric dependencies, the disaggregated spine further reduces architectural coupling and aligns both hardware and software evolution with open, standards-driven ecosystems.

Reference Implementation

The disaggregated spine architecture described above can be implemented using Nexthop AI's scale-across spine and scale-across leaf platforms.

The NH-5010 implements the scale-across spine role, supporting Data Center Interconnect requirements including MACsec encryption, 400 GbE and 800 GbE ZR optical module support, expanded routing scale, and deep packet buffering.

The NH-4010 implements the scale-across leaf role, optimized for intra–data center traffic with high port density, low latency, and efficient aggregation within the fabric-facing tier.

Additional technical specifications are available in the respective product datasheets:

- NH-5010 Datasheet
- NH-4010 Datasheet

References

1. Nexthop Learning Hub for Disaggregated Spine architecture
2. SONiC working group for Disaggregated Spine,
3. OCP 2025 update on Disaggregated Spine architecture and deployment